Chapter 9

Human, Not Humanoid, Robots

Domenico Parisi

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.70117

Abstract

Robots that resemble human beings can be useful artefacts (humanoid robots) or they can be a new way of expressing scientific theories about human beings and human societies (human robots), and while humanoid robots must necessarily be physically realized, human robots may be just simulated in a computer. If the simulated robots do everything that human beings do, the theory which has been used to construct the robots explains human behaviour and human societies. This chapter is dedicated to human robots and it describes a number of individual and social human phenomena that have already been replicated by constructing simulated human robots and simulated robotic societies. At the end of the chapter, we briefly discuss some of the problems that human robots will pose to human beings.

Keywords: human robots, humanoid robots, science, technology

1. Human robots as a new science of human beings

"Robot" is an ambiguous word. It has two different meanings. Robots can be physical artefacts with practical applications and economic value or they can be a new science of human beings. For robots as practical artefacts, success is that there are people who are disposed to spend their money to buy them. For robots as science, success is to construct robots that do *everything* that human beings do, because only if scientists are able to construct robots that do *everything* that human beings do, they will finally understand human beings. Since practically useful artefacts that look like human beings and do some of the things that human beings do are called "humanoid" robots, to make the distinction explicit, we will call robots as science "human" robots. In this chapter, we will be concerned with human, not humanoid, robots.

While humanoid robots necessarily are physical robots, human robots may be just simulated in a computer. In fact, human robots are based on the general assumption that the best way

for science to understand X is to simulate X in a computer. If, when the simulation runs in the computer, its results correspond to what scientists empirically know about X, they are entitled to conclude that the theory incorporated in the computer program captures the mechanisms and processes which underlie X and, therefore, it explains X. Computers can be useful to science in many other ways but they are a true scientific revolution if they are used to express scientific theories in a novel way. So far, scientific theories have been expressed using mathematical symbols or using words. Physicists express their theories using mathematical symbols. Scientists who study human beings express their theories by using words—the only exception is economics but economics is not a science but an applied discipline—and words are a problem for science because they tend to have unclear meanings and to mean different things to different scientists—and defining one word by using other words clearly does not solve the problem. The consequence of expressing scientific theories by using words is that scientists rarely agree on the empirical predictions that can be derived from a theory and they spend most of their time to do endless discussions which resemble more those of philosophers than those of scientists. Human beings are more complicated and more difficult to study than nature, but it is the fact that scientists who study human beings express their theories by using words, which is the real reason why the sciences that study human beings and human societies are so much less advanced than the sciences that study nature.

Computer simulations solve the problem. A theory is formulated as a computer program and, when the program runs in the computer, it generates a large number of quantitative results, which are the predictions derived from the theory/simulation. If these results correspond to what scientists empirically know about reality, the theory/simulation is confirmed. If they do not correspond, the theory/simulation must be modified or abandoned.

Human robots are computational theories of human beings. To understand human beings, scientists must construct simulated human beings which are like real human beings and which do all that real human beings do. Humanoid robots reproduce only an extremely limited number of things that human beings do. They have a body which has some external resemblance to the human body, they walk on two legs, they reach and grasp objects with their hands, they express emotions with their face which they do not really feel, and they produce words which they do not really understand. Human robots must have a body which does not only have a human-like external form but also contains internal organs and systems that simulate the internal organs and systems of the human body—not only the brain, but also the heart, the lungs, and the visceral, endocrine, and immune systems [31]. They must be the result of a process of evolution that takes place in a succession of generations and of a process of development and learning that takes place during the course of a robot's life [1, 2]. Each individual robot must be different from all other robots and the robots must have all sorts of pathologies both of the body and of the mind. They must have their own independent motivations and their behaviour must be determined by their motivations. They must actually feel the emotions which they express with their face, their voice, and their body. They must talk to other robots by producing sounds that they actually understand, and they must also talk to themselves without producing audible sounds (think) [25]. They must respond to stimuli which do not arrive to the brain from the external environment or from their own body [27]

http://dx.doi.org/10.5772/intechopen.70117

but are self-generated by their own brain (mental life). They must be born from a female and a male robot, they must live for a certain period of time, and then they must die. And they must be very social. They must live in families, they must have friends, they must cooperate or compete with other robots, and they must organize themselves in societies that have economic and political institutions and that change historically. They must learn by imitating other robots and develop cultures that may change from one generation to the next. They must modify the environment in which they live, they must make artistic artefacts and must expose themselves to the artistic artefacts made by other robots, and they must have religion, philosophy, and science.

But human robots are not only theories. Scientists can also do experiments on their simulated human beings. They can vary the value of the different variables and see the consequences of these variations. Laboratory experiments are a very important scientific tool but, while they are a perfect tool in the hands of physicists, chemists, and biologists, they have many limitations when they are used to study human beings and human societies. Most of what psychologists know about human beings is derived from laboratory experiments but laboratory experiments provide them with a very limited knowledge of human behaviour [3]. First, human behaviour is the result of the interactions of human beings with their environment, but the laboratory is a very simplified environment, which is very different from the real environment. Therefore, what human beings do in an experimental laboratory may be very different from what they do in their real life. Second, outside the experimental laboratory human beings do what they want to do, whereas experimental subjects do what the experimenter wants them to do.

The problem is even more serious for social scientists—anthropologists, sociologists, economists, and political scientists. Social scientists do very few laboratory experiments because the social environment of human beings is almost impossible to reproduce in the laboratory and because social phenomena are very complex and they are much more extended in time than laboratory experiments. Therefore, social scientists are mostly limited to collecting statistical data on the consequences of human behaviour, making interviews, and reading the books of other social scientists.

A robotic science of human beings changes all of this. Since the behaviour of human beings depends on the environment in which they live what observable and measurable aspects of the robots', to understand human behaviour scientists must simulate in the computer not only human beings but also their natural and social environment [4]. And they not only simulated *laboratory* experiments but also simulated *ecological* experiments in which they vary the natural and social environment that their human robots live and see how the robots' behaviour depends on the particular environment in which they live. And they can also do counterfactual experiments. They can let the robots live in an environment which does not exist and see whether the robots behave in the non-existing environment as predicted by their theories.

Expressing scientific theories as computer simulations has another important advantage. Science is divided into disciplines, with some disciplines studying some of the phenomena that make up reality and other disciplines studying other phenomena. The problem is that reality

is not made up of separate classes of phenomena but it is a large ensemble of phenomena which are all connected together and, often, to understand the phenomena studied by one discipline it is necessary to take into account the phenomena studied by other disciplines. Today, there are attempts at addressing this problem by doing what is called *inter-disciplinary* research: scientists of different disciplines discuss and collaborate together to better understand the phenomena which interest them. But inter-disciplinary research does not really solve the problem because it continues to take scientific disciplines as given, with each discipline studying one particular class of phenomena and possessing theories that try to explain that particular class of phenomena.

Both the science of nature and the science of human beings are divided into disciplines but for the science of nature the division into disciplines is not really a problem because physics, chemistry, and biology use the same empirical methods, have very similar conceptual and theoretical traditions, and share a view of nature as made up of physical causes that produce physical effects and as possessing an inherently quantitative character. On the contrary, the division of the science of human beings into disciplines—psychology, anthropology, linguistics, sociology, economics, political science—has very negative consequences because these disciplines do not share the same empirical methods, have very different conceptual and theoretical traditions, and do not have a unified view of the phenomena they study.

Computers change this situation because they make it possible to develop a non-disciplinary science of reality, a science which completely abolishes scientific disciplines. Science is divided into disciplines because scientists are human beings and their brain is too small to formulate theories that take into account and try to explain the data collected by different scientific disciplines. Computers have a much larger and more powerful "brain," with a memory that can contain enormous quantities of data and a computing capacity that can take into account all the relations among the data. For practical reasons, empirical data will continue to be collected by different scientists but the theories-simulations that explain these data and make predictions about them will not be physical, chemical, biological, psychological, or social theories but they will simply be theories of reality.

A robotic science of human beings is a non-disciplinary science that will abolish not only the divisions among the disciplines that study human beings and their societies but also the great division between the disciplines that study nature and those that study human beings and their societies—which is the most serious obstacle to a scientific comprehension of human beings. Clearly, the creation of a non-disciplinary science of human beings will be a gradual process. A robotic science of human beings will begin by constructing robots and societies of robots, which greatly simplify with respect to real human beings and real human societies, but then the robots and the robotic societies will become progressively more complex and more similar to real human beings and real human societies. Today, only some psychologists and some neuroscientists are interested in human robots but human robots will progressively interest, on one side, biologists and chemists and, on the other side, anthropologists, sociologists, economists, political scientists, and even historians.

To realize a complete robotic science of human beings, it is possible to adopt two different strategies. One strategy is based on the principle "one robot/one phenomenon" and adopting

this strategy means to construct different robots each of which reproduces in a more realistic way one single aspect of human behaviour. The other strategy is based on the principle "one robot/many phenomena," and adopting this strategy means to construct one and the same robot that reproduces in a more simplified way many different human phenomena and then to progressively add more details and make the robot more realistic. The second strategy is better than the first one because one and the same human being perceives what is in his or her environment, moves his or her body, remembers, predicts, speaks and understands, thinks, has a variety of motivations and emotions, does things with other human beings, and participate in the creation and functioning of social structures. Therefore, one and the same robot must do all these things.

If this is the final goal of a robotic science of human beings, this science poses a very general and interesting question. Human robots are theories that try to explain human beings by simulating them in a computer, and they are one example of a general principle, which I think in the future will be adopted by all scientists, according to which, whatever phenomenon science wants to explain, what science must do is simulate the phenomenon in the computer. But there is an important difference between scientific theories expressed by using words or mathematical symbols and theories expressed as computer programs. Verbal and mathematical theories *necessarily* simplify with respect to the phenomena they want to understand and their value depends on the goodness of these simplifications. Theories expressed as computer programs begin by reproducing reality in a very simplified way but then scientists can add more and more details until the simulation *completely* replicates reality. What are the consequences of this progressive convergence between scientific theories and reality? When should scientists stop adding more details? I don't know what is the answer to this question, and I wait for suggestions from philosophers of science.

2. Human robots are a non-verbal science of human beings

Human robots pose this and other interesting philosophical problems but understanding human beings by constructing human robots is the opposite of doing philosophy. While philosophy is made of words and of discussion about words, robotics has no use for words. Psychologists and social scientists use words to formulate their theories, and many of these words have a philosophical origin or have been discussed for centuries by philosophers: sensation, perception, attention, memory, thinking, reasoning, planning, motivation, emotion, representation, concept, category, meaning, object, property, action, intention, goal, consciousness, norms, and values. Robotic scientists can use these words only if they can point out what observable and measurable aspects of the robots' behaviour, brain or society they call sensation, perception, attention, memory, motivation, emotion, etc.

Take the word "category," an important word for both psychologists and philosophers. A robot can be said to have categories if it behaves in the *same* way towards *different* things. Here is a very simple example [5]. A population of robots lives in an environment which contains both roundish and angular objects but no two objects have exactly the same shape. The roundish objects are food and the angular objects are poison and, to remain alive and

have offspring, the robot must reach and eat the roundish food objects and avoid the angular poison objects. If, when we look at the robots on the computer screen, we see that the robots approach and eat the roundish objects and avoid the angular objects, we are entitled to say that they possess the category of food and the category of poison because the word “category” is defined not by using other words but by looking at the robots’ behaviour.

And the word “category” can be defined not only by looking at the robots’ behaviour but also by examining the robots’ brain. The robots’ brain is a neural network made of artificial neurons with a level of activation that varies from one cycle of the simulation to the next cycle and of connections between neurons through which one neuron influences the level of activation of another neuron. Each connection has a quantitative weight which can be either a positive number (excitatory connection) or a negative number (inhibitory connection), and it is this weight that determines how the activation level of one neuron influences the activation level of another neuron. The brain of our robots is made of three types of neurons—visual neurons, internal neurons, and motor neurons—and since the visual neurons are connected to the internal neurons and the internal neurons are connected to the motor neurons, what a robot sees determines what the robot does. If we call “pattern of activation” the ensemble of levels of activation of a set of neurons in each cycle the pattern of activation of the visual neurons is caused by the shape of the object that the robot is currently seeing, this pattern of activation causes a pattern of activation in the internal neurons which, in turn, causes a pattern of activation in the motor neurons, and the pattern of activation of the motor neurons causes the robot to approach or avoid the object.

At the beginning of the simulation, the connections of the robots’ neural network have random weights and, therefore, the robots are unable to distinguish between the roundish and the angular objects and to approach the roundish objects and avoid the angular objects. Therefore, on average, these robots do not eat much food and they also eat some poison, which means that they have a short life and are unable to generate many offspring.

The capacity to distinguish between the roundish and the angular objects is acquired through a process that takes place in a succession of generations and simulates biological evolution. The selective reproduction of the robots which, for purely random reasons, have better connection weights in their neural network and, therefore, have some tendency to approach the roundish objects and to avoid the angular objects, and the addition of random changes in the quantitative weights of the connections inherited by the offspring robots from their parent robots (genetic mutations)—which in some cases can result in offspring which are better than their parents—determine, in a succession of generations, the progressive acquisition of the capacity to approach and eat the roundish objects and to avoid the angular objects. Therefore, at the end of the simulation, we can say that the robots have acquired the category of food and the category of poison.

This is what we find when we examine the robots’ behaviour. But we can also ask: What happens in the robots’ brain that make the robots approach and eat the roundish objects and avoid the angular objects? To answer this question, we look at how the different objects are “represented” in the robots’ brain, where the neural “representation” of an object is the pattern of activation of the internal neurons of a robot’s neural network which is caused by the sight of the object. What we find is that while in the robots of the first generation the neural

http://dx.doi.org/10.5772/intechopen.70117

representations of the roundish and angular objects are confused together; after a certain number of generations, the roundish objects cause very similar patterns of activation in the internal neurons and the same for the angular objects, but the patterns of activation caused by the roundish objects are different from the patterns of activation caused by the angular objects. This means that the robots have evolved the capacity to categorize some objects as roundish and other objects as angular.

We have described this simulation to illustrate how a robotic science of human beings treats words. Robotic scientists can use words—in our case, the word "category" and the word "representation"—but only if they can point out to what these words refer to either in the robots' behaviour or in the robots' brain. As we have already said, this is not what happens in the traditional sciences that study human behaviour and human societies. Scientists dedicate much of their time to defining words by using other words and to discussing the meaning of a word without never reaching an agreement. The consequence is that from a verbally formulated theory different scientists may derive different predictions and, therefore, their theories can never be confirmed or disconfirmed by what is empirically observed and measured. By not using words or by using words only if their meaning can be translated in what is observed and quantitatively measured, the robotic science of human beings solves this problem.

3. Only a robotic science of human beings can look at human beings with the detachment required by science

Scientists are human beings and, unlike when they study nature, when they study human beings they are almost inevitably influenced by their values, desires, and fears. Therefore, from a verbally formulated theory, scientists may not only derive different empirical predictions because the theory is unclear and ambiguous but they may also be influenced by their values in choosing which predictions to derive from the theory—which is another reason why the sciences that study human beings and human societies are so much less advanced than the sciences that study nature.

This changes if scientists express their theories of human beings and human societies by constructing human robots and human robotic societies. What the robots do and why they do are under the eyes of everyone and scientists cannot deny the evidence provided by the robots. This is another important advantage, which is provided by a robotic science of human beings and human societies. This science will make it possible to study human beings and human societies with the same detachment with which natural scientists study nature.

A related problem is that scientists belong to different cultures and, while this has no consequences when they study nature and when they study human beings and human societies, they tend to be influenced by their culture. This is very clear for anthropologists but it is a general problem for the sciences that study human beings and human societies because science must be universal and independent from culture. Studying human beings and human societies by simulating them in a computer solves this problem. By constructing robotic societies that have different cultures, scientists will be able to look at human beings and their cultures—including their own culture—with the necessary detachment.

4. Human robots must have their own motivations and they must do they want to do

Although human robots will make it possible for science to know human beings much better than its previous attempts at knowing them, they will also pose many problems to human beings. Robots as technologies already pose problems to human beings but, since these problems are discussed in the other chapters of this book, we will concentrate on the problems that robots as science will pose to human beings.

The most serious of these problems is due to the fact that while humanoid robots are constructed to satisfy *our* motivations, human robots must have *their* own motivations and they must do want *they* want to do, not what *we* want them to do. Some humanoid robots are said to be "autonomous" but, since humanoid robots are technological artefacts, technological artefacts cannot be really autonomous. They can autonomously decide what to do to reach a certain goal but the goal is decided by us. A humanoid robot can autonomously decide how to move its arm and its fingers to reach and grasp an object with its hand but *we* decide that it must reach and grasp the object with its hand. Therefore, humanoid robots can be cognitively (behaviourally) but not motivationally autonomous. Human robots must be both cognitively and motivationally autonomous because human beings are both cognitively and motivationally autonomous. They must decide both that they want to reach and grasp the object with their hand and know how to move their arm and their fingers to reach and grasp the object.

Motivations are the most important component of human behaviour—and of the behaviour of all animals. One often hears that behaviour is caused by stimuli, but this is not true. An individual's behaviour is *guided* by stimuli but it is *caused* by the individual's motivations. The robots described in Section 2 had the motivation to eat food and the motivation not to eat poison, and the real cause of their behaviour was these two motivations. Seeing a roundish object or an angular object only guided them towards the roundish object or away from the angular object.

Since the motivations of those robots were only two and they always had the same strength, it was rather easy for the robots to decide which of the two motivations to satisfy with their behaviour at any given time: seeing a roundish objects activated one motivation and seeing an angular object activated the other motivation. Human beings have a much greater number of different motivations and the strength of these motivations can change from one moment to the next as a function of various factors. Therefore, it is more difficult for human beings to decide which motivation they should try to satisfy with their behaviour at any given time. Their motivations lie dormant in their brain and in their body and they are activated not only by the external stimuli—like the two motivations of the robots of Section 2—but also by stimuli self-generated inside their brain and inside their body. The problem is that human beings—and all animals—cannot satisfy two or more motivations at the same time and, therefore, in any given moment, they must decide which of their different motivations they should try to satisfy with their behaviour. Since their motivations have different strengths and this strength varies with the circumstances, they try to satisfy the motivation which at any given time has the greatest strength.

http://dx.doi.org/10.5772/intechopen.70117

This is a simple example of robots that have two motivations whose strength varies from time to time [6]. The robots need both energy and water to remain alive and, since their body constantly consumes both energy and water, they must both eat food (green objects) and drink water (white objects). The robots' body contains two internal stores, one for energy and the other one for water, and the robots' brain has two additional sets of sensory neurons whose activation level reflects the quantity of energy and the quantity of water currently contained in the two bodily stores. These neurons are activated when the quantity of energy or water contained in the robots' body is below a certain level and it is their activation that makes the robots feel hungry or thirsty. The capacity of the robots to respond to hunger by looking for food and to thirst by looking for water evolves in a succession of generations. At the beginning of the simulation, the robots do not look for the green objects when they feel hungry and for the white objects when they feel thirsty but, after a certain number of generations, the robots look for food and ignore water when there is little energy in their body and they feel hungry and they look for water and ignore food if there is little water in their body and they feel thirsty.

Although motivations, not external stimuli, are the real causes of behaviour, external stimuli have an important motivational role because they may activate different motivations. For example, the sight of a predator may activate in a robot the motivation to fly away from the predator while the sight of a robot of the opposite sex may activate the motivation to mate with the robot of the opposite sex. This is true for both animal robots and human robots. But human robots must be more complex because their motivations must be activated not only by external stimuli (the sight of a predator robot or the sight of a robot of the opposite sex) or by internal stimuli self-generated by their own body (hunger and thirst) but also by internal stimuli self-generated by their own brain (thoughts, memories, and imaginations).

But human robots must not only have their own motivations. They must also feel emotions [32] because emotions are a submechanism of motivations [7]. Emotions are states/processes of the body/brain that increase the current strength of one particular motivation so that the individual will choose to satisfy this motivation rather than other motivations. Robots which feel emotions are robots whose brain includes a set of neurons that function differently from the other neurons. First, when they are activated, their activation persists for a certain number of input/output cycles and, second, they send stimuli to other organs and systems that are inside the body such as the heart and the visceral system [31] and these other organs and systems respond by sending stimuli to the brain which modify the strength of the various motivations. This emotional circuit makes the motivational choices of the robot more adaptive—although they may also cause psychical disturbances, for example, if the robot finds it impossible to satisfy a motivation which, for the robot, has a very high strength.

Here is one example of how emotions can help robots to take better motivational decisions [8]. The robots live in an environment which not only contains food objects that they must eat to remain alive but also contains a predator that can suddenly appear and kill the robots. For adaptive reasons, the motivation to fly away from the predator is intrinsically stronger than the motivation to eat and, in fact, when the predator appears, the robots cease to look for food and they fly away from the predator. We compare two populations of robots. The neural network of the robots of one population has only sensory neurons for food and sensory neurons for the predator, whereas the neural network of the robots of the other population, in

addition to these sensory neurons, has a set of emotional neurons. These emotional neurons are not activated by the sight of food but they are only activated by the sight of the predator, and their activation persists even if the robot flies away and, therefore, it ceases to see the predator. Since these emotional neurons send their connections to the motor neurons, they influence the robots' behaviour.

When we compare the two populations of robots, we find that the robots with the emotional neurons are less likely to be killed by the predator compared to the robots without the emotional neurons. If we look at the robots' behaviour on the computer screen, we see that the robots with the emotional neurons immediately run away from the predator as soon as they see the predator and they continue to run away even if they cease to see the predator, whereas the robots without the emotional neurons are less good at flying away and, therefore, they are more easily killed by the predator. The robots with the emotional neurons in their neural network can be said to experience the emotion of fear, and experiencing the emotion of fear helps them to remain alive.

Here is another example that demonstrates how feeling emotions helps the robots to take better motivational decisions. The robots we have described so far do not have a sex and they do not need a mate to generate offspring. The new robots are males and females, and to generate offspring, a robot must mate with a robot of the other sex. (The male robots look differently from the female robots.) This means that these robots also have two motivations to satisfy, the motivation to eat to remain alive and the motivation to mate to have offspring, and they must divide their time between looking for food and looking for a robot of the opposite sex. Again, we compare a population of robots with a set of emotional neurons in their brain and another population of robots without emotional neurons. The results are that the robots with the emotional neurons in their brain are more attracted by the robots of the opposite sex and, therefore, they have more offspring than the robots without the emotional neurons. They eat what is sufficient to remain alive but, unlike the robots without the emotional neurons, as soon as they see a robot of the opposite sex, they ignore food and approach the robot of the opposite sex. Unlike the robots without the emotional neurons, they can be said to experience the emotion of "sexual attraction."

Like motivations, emotions clearly distinguish between robots as science and robots as technology, between human and humanoid robots. Some of today's humanoid robots *express* emotions with the movements of their face or with the tone of their voice because this makes them more attractive for potential buyers, but they do not really feel these emotions. Theirs are *unfelt* emotions—an obvious contradiction. On the contrary, human robots must actually "feel" emotions because human beings actually feel emotions, and they must express their emotions with their face, voice, and body but also keep their emotions for themselves because this what human beings do.

Robots that have their own motivations and emotions contradict Asimov's three laws of robotics. They must do what they want to do because human beings do what they want to do and they cannot obey laws unless they themselves promulgate these laws because human beings obey (most of the times) laws that they themselves have promulgated. In fact, human robots are not really robots if the word "robot" must continue to have its original meaning of "slave worker," because human beings are not slave workers.

5. Human robots must be social robots

Another characteristic of human robots that will pose problems to human beings is that human robots will need to be very social robots because human beings are very social animals. Human beings live with other human beings, they spend most of their life doing things with other human beings, they have cultures that make them behave and think like some other beings but unlike other human beings, and they have economic and political institutions. Therefore, human robots must live and interact with other robots, they must talk with other robots, they must live in societies that are like human societies, and they must develop cultures.

Although today one often hears of social robots, social robots are not really social because they interact with us, not between them—and the reason is obvious. Today's "social" robots are constructed to take care of old or ill human beings, to entertain human beings of all ages, and to do other things with human beings because this is what makes it possible to sell them and produce profits. But they do not interact with other robots. The only robots which interact with other robots are those of swarm robotics but the robots of swarm robotics not only resemble much simpler animals than human beings but the robots that make up a swarm of robots are all identical and for them success is only collective success, while no two human beings—and no two members of the any animal species—are identical and a crucial factor in social life is the contrast between individual and collective success.

In fact, a robotic social science that lets us better understand the enormous variety of human social phenomena still does not exist. Today, some human social phenomena are simulated in the computer by using "agents," not robots. Agents do not have a body, do not have a brain, and they do not live in a physical environment. They receive abstract inputs from other agents and, on the basis of very simple rules, they respond by sending abstract inputs to other agents. Agent-based social simulations are useful tools but they must be seen as only a first step towards a robotic social science. If we want to really understand human social behaviour, we must replace agents with robots because human beings do not cease to have a body and a brain and to live in a physical environment when they interact with other human beings and create societies and cultures [9–12, 23, 28].

In this section, we describe robots that simulate some very basic aspects of human sociality but, since human sociality is very complex, most of the work remains to be done.

A very important aspect of human social behaviour is language. Human beings interact together by using language and, therefore, human robots must have language. Humanoid robots seem to understand the linguistic sounds that they hear and the linguistic sounds that they themselves produce but this is not really true. They are only programmed to respond in specific ways to specific sounds and to produce specific sounds in the appropriate circumstances. To have language is something different. It is to possess a neural network which, in addition to sensory and motor neurons, has two sets of reciprocally connected internal neurons. The patterns of activation of the first set of internal neurons are the neural representations of the different objects which the robot sees, whereas the patterns of activation of the second set of internal neurons are the neural representation of the different sounds which the robot hears. The robot learns language in a succession of trials and, at the end of learning, since the two

sets of internal neurons are reciprocally connected, seeing an object causes the appearance of the neural representation not only of the object but also of the sound that designates the object (speaking) and hearing a sound causes the neural representation not only of the sound but also of the object which is designated by the sound (understanding) [18, 19, 26, 30].

What difference does it make to have language? To answer this question, we return to the robots we have described in Section 2. To remain alive and reproduce, those robots had to distinguish between two categories of objects, roundish (food) and angular objects (poison), and to eat the first category of objects and avoid the second category of objects. Now we teach these robots to understand language and we find that if during their life these robots learn to respond to one sound ("food") by approaching and eating the roundish objects they see and to respond to a different sound ("poison") by avoiding the angular objects they see, they live a longer life and have more offspring. Why? If we examine the neural networks of the robots, we find that the neural representation of the roundish object is more similar than they were for the robots without language and the same for the neural representation of the angular objects. Language makes behaviour more effective.

Of course, language has many other uses and many other aspects. We have constructed robots that illustrate some of these other uses and aspects [24, 25] but, again, most of the work is still to be done.

We now turn to other aspects of human sociality and we begin by describing robots which, like human beings, are males and females and, to reproduce, must mate with a robot of the other sex [13]. Male and female robots have different colours and this makes them recognizable as males or females by the other robots. But the real difference between male and female robots is that, after mating with a female robot, a male robot can immediately reproductively mate with another female robot and generate other offspring, whereas female robots have a period during which they are non-reproductive due to pregnancy, hormonal changes, lactation, and other factors and also their colour changes so that males can distinguish them from non-pregnant females. Both male and female robots do not have only the motivation to mate and have offspring but they also have the motivation to eat because if they don't eat, they die. The question is: What motivation is stronger, mating or eating?

The answer depends on the sex of the robots. At the end of the simulation, we bring the robots, one at a time, into an experimental laboratory and we let them choose between two alternatives. The results are the following. If male robots must choose between a piece of food and a reproductive female, almost all male robots prefer the non-reproductive female to the piece of food. Why? The answer is that, while in the robots' environment food is always available, this is not true for reproductive females because at any given time many female robots are non-reproductive. Therefore, unless they are very hungry, male robots are more interested in reproductive females than in food. On the contrary, if male robots must choose between food and a non-reproductive female, they almost completely ignore the non-reproductive female and they choose food.

Female robots do not only behave differently from male robots but they also behave differently when they are reproductive and when they are non-reproductive. If reproductive females must choose between food and a male robot, they tend to choose food rather that the

male robot, and this implies a strategy of using one's time to look for food and simply waiting for a male to mate with because males are always looking for non-pregnant females. But what is interesting is that the same happens if a non-pregnant female must choose between a male and another non-pregnant female. The non-pregnant female prefers the non-pregnant female to the male. Why? Perhaps because, in the real environment, staying close to other non-pregnant females makes non-pregnant females more attractive for males. Ignoring males is even more frequent among non-reproductive females. A non-reproductive female must choose between a male and food or between a non-reproductive female and food, almost always chooses food.

The next step is families. Families are groups of genetically related individuals who live together and, since families are a very important human social phenomenon, human robots must live in families. The members of a family—mother, father, daughters, sons, grandmothers, grandfathers—live together because by living together they can help each other, and they are motivated to help each other because this increases the probability that their genes or the copy of their genes possessed by their relatives will remain in the genetic pool of the population (kin-selection).

We have simulated some simple phenomena concerning human families. In one simulation, when they are very young and therefore they are still unable to find the food which exists in the environment, the robots evolve the behaviour to follow their parents rather than other robots because, in parallel, parents have evolved the behaviour of feeding their very young offspring. In another simulation, sisters and brothers evolve the behaviour of giving some of their food to their sisters and brothers but not to extraneous robots and, in a third simulation, grandmothers and grandfathers evolve the behaviour of feeding their nephews even if this may cost them their life.

Other social phenomena go beyond families and concern entire communities. Social proximity is (or was) a pre-condition for social interaction and it may be influenced by the nature of the environment. Consider two environments. In one environment, food exists in all parts of the environment, whereas in the other environment food only exists in certain parts of the environment. What we find is that while the robots of the first environment do not live near to one another, the robots of the second environment live together in communities in those parts of the environment that contains food [14]. But robots may live near to one another independently of the nature of the environment because, if they live near to one another, they may coordinate their behaviour and display useful collective behaviours [15, 16].

Human beings can live in smaller or larger communities and human history is characterized by the progressive increase in the size of human communities to the point that, today, human beings tend to live in a single global community. To reproduce this phenomenon, we compare two populations of robots both living in a seasonal environment. The robots of one population are divided into a certain number of small communities, each living in its small territory, whereas the robots of the other population are a single community and their territory is the entire environment. The results of the simulation are that the robots that form a single large community and go everywhere in the environment looking for food continue to exist, whereas the robots that are divided into small communities become extinct.

The robots I have described so far need only one type of food to remain alive. However, if to remain alive the robots need to eat two different types of food and the two types of food are in two different parts of the environment, the robots must continuously move from one to the other part of the environment, and this is very expensive in terms of both time and energy. In these circumstances, the robots spontaneously evolve the exchange of food. Some robots tend to live in the part of the environment which contains one type of food and other robots in the part of the environment that contains the other type of food, and then the robots meet together to exchange one type of food for the other type of food [22, 29].

Food is only one type of good, where a good is anything that human beings try with their behaviour to have. Human beings want to have many different goods because their goods are not only those that exist in nature but they produce always new goods by using the existing goods: clothes, homes, tools, cars, and many other things. The increase in the number of goods that human beings want to have has caused the invention of money. The invention of money can be simulated in the following way ([17], Chapter 11). We begin with a population of robots that want to have many different goods and, since a robot cannot produce all these goods, the robots must meet together to exchange their goods. But when two robots meet together to exchange their goods, one or both robots may not need the particular goods that the other robot has and, therefore, the exchange cannot take place. To solve this problem, the robots spontaneously invent money. At the end of the simulation, we find that one particular good is exchanged in all exchanges, and this good is money. All the robots want to have money because they can obtain all sort of goods from other robots in exchange for money. The exchange of goods has become buying and selling.

We conclude this section by mentioning two general characteristics of social behaviour which still need to be reproduced with human robots.

The social environment and the natural environment are very different environments and what human robots must do to obtain what they want from the two environments is very different [21]. To obtain what they want from the natural environment, they must simply act physically on the natural environment. To obtain what they want from another robot, they must change the other robot's brain. And if we ask what they must change in the other robot's brain, the answer is: its motivations. As we have seen in Section 4, what human robots do depends on their motivations and on the current strength of their motivations, and their behaviour is aimed at satisfying the motivation which currently has the greatest strength. Therefore, to obtain what it wants from another robot, a robot must change the current strength of the other robot's motivations. This is social behaviour: changing the motivations of others so that they do what one wants them to do. To change the motivations of other robots, a robot can send all sorts of sensory inputs to their other robots' brain. It can talk to them, it can modify its external physical appearance by dressing and by decorating its body, and it can express its emotions with its face, its voice, and its body.

The social environment has other characteristics which make it different from the natural environment. An important capacity of human beings is the capacity to predict the consequences of their behaviour and to decide to actually execute the behaviour only if they consider these consequences as good [6]. This capacity can be simulated with robots in the following way

[20]. The neural network of the robots has two additional set of internal neurons, the prediction neurons and the "good/bad" neurons. The predictions neurons are activated by the current sensory input and by a planned but still not physically executed behaviour in response to the current sensory input. The "good/bad" neurons are activated by the prediction neurons. When the robots' neural network receives a sensory input from the environment, it does not automatically responds to this input by executing some behaviour but it plans some behaviour, predicts its consequences, judges if these consequences are good or bad, and physically executes the behaviour only if they are good.

But there is an important difference between predicting the effects of one's behaviour on the natural environment and on the social environment. To predict what will happen in the natural environment, human beings must take into considerations only the sensory inputs which arrive to their sensory organs from the natural environment. To predict what another individual will do, they must take into consideration not only the sensory input which currently arrives to their sensory organs from the other individual but also the sensory input which currently arrives to the other individual's sensory organs and what are the other individual's motivations. And there is also another problem. Human beings are more different from one another than inanimate objects and this makes their behaviour more difficult to predict. Inanimate objects obey more or less the same laws and these are relatively simple laws which are not so difficult to discover. The behaviour of human beings does not only obey more complex laws but each individual is so different from all other individuals that his or her behaviour cannot be predicted by only using general laws.

6. What problems will human robots pose to human beings?

As we have already said, human robots will pose problems to human beings and these problems will be more serious than the problems posed to human beings by humanoid robots or, more generally, by robots. This last section is dedicated to a very brief discussion of these problems.

As we have already said, while humanoid robots must necessarily be physically realized to be useful to those who buy them, human robots may be useful to science even if they are only simulated in a computer. However, in the future, human robots will also be physically realized and, when this will happen, they will pose more serious problems to human beings than physically realized humanoid robots.

However, human robots can pose problems to human beings even if they are only simulated in a computer. Computers are interactive devices and, therefore, human beings will have the possibility to interact with the simulated human robots which are inside their computer. Today, many human beings—especially young human beings—spend much of their time by interacting with the digital environment rather than with the real environment. But when the digital environment will contain human robots, it is possible that a much greater number of human beings—of all ages—will prefer to live in a simulated social environment made of simulated human beings rather than in the real social environment made of real human beings.

And the simulated human robots may convince them to do what is not in their interests or they may want to damage them in other ways. One might object that human beings can always switch off the computer but the simulated human robots may convince them not to do so.

Can we control human robots so that they do not do what *we* don't want them to do? This is already a problem for robots as technologies but for robots as technologies the problem can be solved by emanating laws that prohibit the construction of certain types of robots. This cannot be done with robots as science. We can put limits to technology, but can we put limits to science?

Human robots—whether simulated in a computer or physically realized—may also pose embarrassing questions to human beings. If someone constructs a robot which is like my friend Gabriele, who is Gabriele, my friend or the robot? To really understand me, I must construct a robot which is like me? Who am I, I or a robot which is like me?

But the true danger for human beings of a robotic science of human beings is that it will let human beings know themselves as science knows nature. Human robots will not only demonstrate that human beings are only nature but they will project the cold light of science on everything that we are, do, think, and feel. According to the Greek philosopher Democritus, "truth lies in the abyss." Human robots will let human beings fall in the abyss.

7. Conclusion

Unlike humanoid robots that are practically useful physical artefacts which have some resemblances to human beings, human robots are computer simulations that must reproduce everything that human beings are and do and will make it possible for science to finally understand human beings and their societies as it understands nature. In this chapter, we have described a number of individual and social human phenomena which have already been reproduced by constructing human robots and societies of human robots and we have briefly discussed some of the very serious problems that human robots will pose to human beings.

Author details

Domenico Parisi

Address all correspondence to: domenico.parisi@istc.cnr.it

Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

References

[1] Nolfi S, Elman JL, Parisi D. Learning and evolution in neural networks. Adaptive Behaviour. 1994;**3**:5-28. DOI: 10.1177/105971239400300102

[2] Elman JL, Bates EA, Johnson MA, Karmiloff-Smith A, Parisi D, Plunkett K. Rethinking Innateness. A Connectionist Perspective on Development. Cambridge, MA: MIT Press; 1996

[3] Heinrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Brain and Behavioural Science. 2010;**33**:61-135

[4] Duchon AP, Kaebling LP, Warren WH. Ecological robotics. Adaptive Behaviour. 1998;**6**: 473-507. DOI: 1177/105971239800600306

[5] Mirolli M, Parisi D. How can we explain the emergence of a language which benefits the hearer and not the speaker? Connection Science. 2005;**17**:307-324. DOI: 10.1080/0954 0090500177539

[6] Cecconi F, Parisi D. Neural networks with motivational units. In: From Animals to Animats 2. Cambridge, MA: MIT Press; 1992. pp. 167-181

[7] Arbib MA and Fellows JM. Emotions. From brain to robots. Trends in Cognitive Science. 2004;**6**:554-561. DOI: 10.1016/j.tics.2004.10.004

[8] Parisi D, Petrosino G. Robots that *have* emotions. Adaptive Behaviour. 2010;**18**:453-469. DOI: 10.1177/1059712310388528

[9] Denaro D, Parisi D. Cultural evolution in a population of neural networks. In: Marinaro M, Tagliaferri R, editors, Neural Nets. London: Springer; 1996

[10] Ugolini M, Parisi D. Simulating the evolution of artefacts. In Floreano F, Nicoud J-D, Mondada F, editors. Advances in Artificial Life. London: Springer; 1999. pp. 489-498. DOI:10.1007/3-540-48304-7_67

[11] Acerbi A, Parisi, D. Cultural transmission between and within generations. Journal of Artificial Societies and Social Simulation. 2006;**9**(1)

[12] Parisi D, Nolfi S. Sociality in embodied social agents. In Run R, editor. Cognition and Multi-Agent Interactions. Cambridge: Cambridge University Press; 2006

[13] Da Rold E, Petrosino G, Parisi D. Male and female robots. Adaptive Behaviour. 2011;**2**: 5-407. DOI: 1177/1059712311417737

[14] Cecconi F, Denaro D, Parisi D. Social aggregation in evolving neural networks. In: Castelfranchi C, Werner E, editors. Artificial Social Systems. London: Springer; 1994. pp. 41-54. DOI: 10.1007/3-540.58266-5_3

[15] Baldassarre G, Parisi D, Nolfi S. Distributed coordination of simulated robots based on self-organization. Artificial Life. 2006;**12**:298-311. DOI: 10.1162/artl.2006.12.3.289

[16] Baldassarre G, Nolfi S, Parisi D. Evolving mobile robots able to display collective behaviour. Artificial Life. 2003;**9**:255-267. DOI: 10.1162/106454603322392460

[17] Parisi D. Future Robots. Towards a Robotic Science of Human Beings. Amsterdam: John Benjamins; 2014

[18] Cangelosi D, Parisi D. The emergence of a 'language' in an evolving population of neural networks. Connection Science. 1998;**10**:83-97. DOI: 10.1080/0954000998116512

[19] Cangelosi A, Parisi D. The processing the verbs and nouns in neural networks: Insights from synthetic bran imaging. Brain and Language. 2004;**2**:401-408. DOI: 10/1016/S0093-934X(03)00353-3

[20] Cecconi F, Parisi D. Learning to predict the consequences of actions. In: Eckmiller R, Hartmann G, Hauske G, editors. Parallel Processing in Neural Systems and Computers. Amsterdam: Elsevier; 1990. pp. 237-240

[21] Cecconi F, Parisi D. Individual versus social survival strategies. Journal of Artificial Societies and Social Simulation. 1998;**1**(2)

[22] Delre SA, Parisi D. Information and cooperation in a simulated labour market. A computational model of the evolution of workers and firms. In: Salzano M, Colander D, editors. Complexity Hints for Economic Policy. London: Springer; 2007

[23] Gigliotta O, Miglino O, Parisi D. Groups of agents with a leader. Journal of Artificial Societies and Social Simulation. 2007;**10**:1-10

[24] Mirolli M, Parisi D. Language, altruism, and docility: How cultural learning can favour language evolution. In: Pollock JB, Bedau M, Husbands P, Ikegami T, Watson RA, editors, Artificial Life. Vol. 9. Cambridge, MA: MIT Press; 2004. pp. 182-187

[25] 25. Mirolli M, Parisi D. Talking to oneself as a selective pressure for the emergence of language. In: Cangelosi A, Smith ADM, Smith K, editors. Proceedings of the 6th International Conference on the Evolution of Language. Singapore: World Scientific; 2006. pp. 182-187

[26] Mirolli M, Cecconi D, Parisi D. A neural network model for explaining the asymmetries between linguistic production and linguistic comprehension. In: Vosniadou S, Kaiser D, Protopapas A, editors. Proceedings of the European Cognitive Science Conference. Hillsdale, NJ: Erlbaum; 2007. pp. 670-675

[27] Parisi D, Cecconi F, Nolfi S. Econets: Neural networks that learn in an environment. Network. 1990;**1**:149-168. DOI: 10.1088/0954-898X/1/2003

[28] Parisi D. Cultural evolution in neural networks. IEEExpert. 1997;**12**:9-11. DOI:10.1109/64.608170

[29] Parisi D. What to do with a surplus. In: Conte R, Hegselman R, Terna P, editors. Simulating Social Phenomena. London: Springer; 1997. pp. 133-151. DOI: 10.1007./978-3-662-03366-1_10

[30] Parisi D. An artificial life approach to language. Brain and Language. 1997;**59**:121-146. DOI: 10.1006/brln.1997-1815

[31] Parisi D. Internal robotics. Connection Science. 2004;**16**:325-338. DOI: 10.1080/0954009412331314768. (328-354)

[32] Ziemke T. The role of emotions in biological and robotic autonomy. Biosystems. 2008;**91**:401-408. DOI: 10.1016/.biosystems.2007.15.015